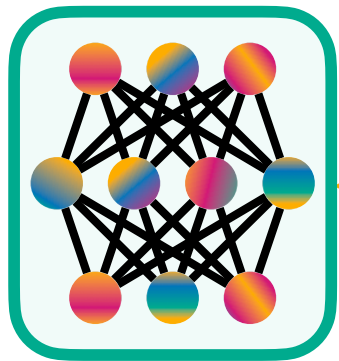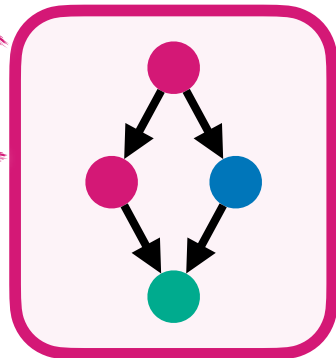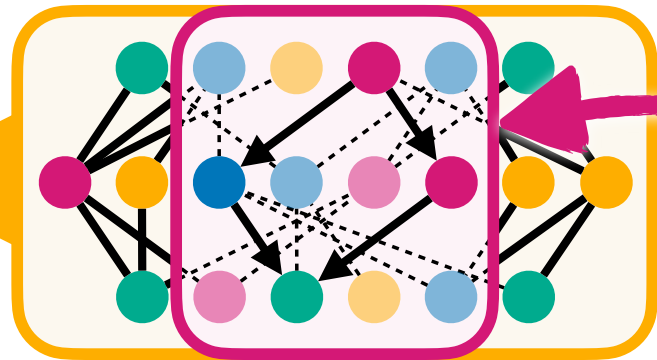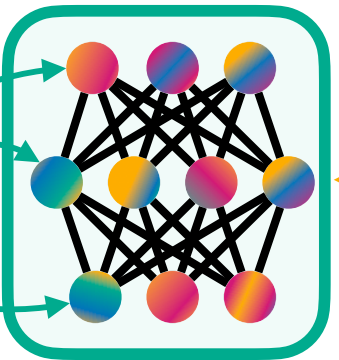**Observed models trained on similar task and data**
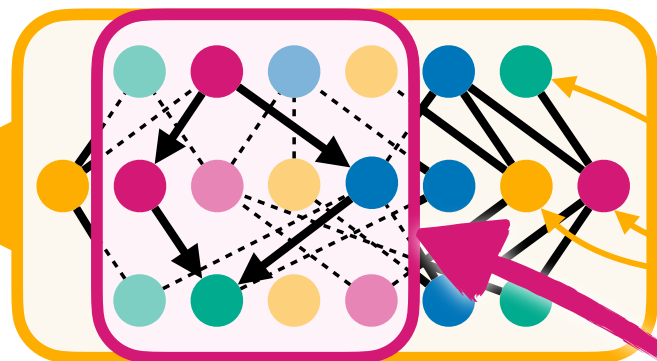
Hypothetical disentangled models

Features

Neurons

Motif